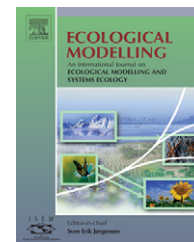


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/ecolmodel

An improved state-parameter analysis of ecosystem models using data assimilation

M. Chen^{a,*}, S. Liu^a, L.L. Tieszen^b, D.Y. Hollinger^c

^a SAIC, USGS Center for Earth Resources Observation and Science (EROS), Sioux Falls, SD 57198 USA

^b USGS Center for Earth Resources Observation and Science, Sioux Falls, SD 57198, USA

^c USDA Forest Service, Northern Research Station, Durham, NH 03824, USA

ARTICLE INFO

Article history:

Published on line 31 August 2008

Keywords:

Gross primary production

Ecosystem respiration

Net ecosystem exchange

Smoothed ensemble Kalman filter

AmeriFlux data

ABSTRACT

Much of the effort spent in developing data assimilation methods for carbon dynamics analysis has focused on estimating optimal values for either model parameters or state variables. The main weakness of estimating parameter values alone (i.e., without considering state variables) is that all errors from input, output, and model structure are attributed to model parameter uncertainties. On the other hand, the accuracy of estimating state variables may be lowered if the temporal evolution of parameter values is not incorporated. This research develops a smoothed ensemble Kalman filter (SEnKF) by combining ensemble Kalman filter with kernel smoothing technique. SEnKF has following characteristics: (1) to estimate simultaneously the model states and parameters through concatenating unknown parameters and state variables into a joint state vector; (2) to mitigate dramatic, sudden changes of parameter values in parameter sampling and parameter evolution process, and control narrowing of parameter variance which results in filter divergence through adjusting smoothing factor in kernel smoothing algorithm; (3) to assimilate recursively data into the model and thus detect possible time variation of parameters; and (4) to address properly various sources of uncertainties stemming from input, output and parameter uncertainties. The SEnKF is tested by assimilating observed fluxes of carbon dioxide and environmental driving factor data from an AmeriFlux forest station located near Howland, Maine, USA, into a partition eddy flux model. Our analysis demonstrates that model parameters, such as light use efficiency, respiration coefficients, minimum and optimum temperatures for photosynthetic activity, and others, are highly constrained by eddy flux data at daily-to-seasonal time scales. The SEnKF stabilizes parameter values quickly regardless of the initial values of the parameters. Potential ecosystem light use efficiency demonstrates a strong seasonality. Results show that the simultaneous parameter estimation procedure significantly improves model predictions. Results also show that the SEnKF can dramatically reduce the variance in state variables stemming from the uncertainty of parameters and driving variables. The SEnKF is a robust and effective algorithm in evaluating and developing ecosystem models and in improving the understanding and quantification of carbon cycle parameters and processes.

© 2008 Elsevier B.V. All rights reserved.

* Corresponding author at: 47914 252nd Street, Sioux Falls, SD 57198-001, USA. Tel.: +1 605 594 6545; fax: +1 605 594 6529.

E-mail address: mchen@usgs.gov (M. Chen).

0304-3800/\$ – see front matter © 2008 Elsevier B.V. All rights reserved.

doi:10.1016/j.ecolmodel.2008.07.013

1. Introduction

There are inherent limitations in the measurement and modelling of ecosystem carbon dynamics. Measurements are usually patchy in space and discontinuous in time, and modelling of carbon dynamics is always built on a set of principles coupled with assumptions and imperfectly defined parameters. Advanced data assimilation techniques based on statistics or optimization theory can mitigate these limitations by combining a series of measurements with dynamic models. To date, much of the effort spent in developing data assimilation methods for carbon dynamics analysis has focused on estimating optimal values for either model parameters (e.g., [Braswell et al., 2005](#); [Knorr and Kattge, 2005](#); [Wang et al., 2006](#)) or state variables ([Bond-Lamberty et al., 2005](#)), but only rarely both ([Gove and Hollinger, 2006](#)). Methods that focus on estimating parameter values alone (i.e., without considering state variables) generally attempt to minimize long-term prediction error by using a historical batch of data that assumes time-invariant parameters. The procedures used to process the historical data as a whole lack the flexibility to investigate the possibility that model parameters change over time. Although there have been some attempts to partition data into a number of subsets in time order, such partitions are inevitably subjective ([Reichstein et al., 2005](#)). Their main weakness is that all errors from input, output, and model structure are attributed solely to model parameter uncertainties. Sequential data assimilation procedures such as the ensemble Kalman filter (EnKF) have the potential to overcome this drawback by explicitly taking all sources of uncertainty into account ([Evensen, 2003](#); [Nichol et al., 2002](#)). However, the successful application of the EnKF focuses primarily on estimating time-varying state variables under the typical presumption that the parameters are to be specified in advance. For example, [Williams et al. \(2005\)](#) successfully used the EnKF to improve analysis of forest carbon dynamics in a young ponderosa pine stand. Because the ecosystems are too complex to guarantee that the model can thoroughly represent the state and change of the system, it is incisive to diagnose the adequacy of model structure and adjust the model through time variation of parameters together with state variables. Therefore it is required to develop a novel sequential data assimilation procedure that will provide simultaneous estimation of time-varying model states and parameters.

Recently there have been a few of studies with encouraging results concerning ensemble-based state-parameter estimation in atmospheric science and hydrology. Such as, [Anderson \(2001\)](#) applied the EnKF to 40-parameter Lorenz model to estimate simultaneously state variables and forcing parameters ([Lorenz, 1996](#)). [Annan et al. \(2005\)](#) had further investigated the applicability of the EnKF to an intermediate complexity general circulation model (GCM), in which the initial-condition sensitivity of model behaviour was naturally minimized by climatological response of the atmosphere. [Aksoy et al. \(2006\)](#) used the EnKF to perform simultaneous state-parameter estimation of the sea-breeze model in the presence of initial-condition error. In their studies, the state-parameter estimation was to use EnKF on a joint vector concatenated uncertain parameters and state variables,

and coin covariance inflation to reduce the effect of filter divergence caused by narrowing of parameter variances through repeated cycling of the EnKF. [Moradkhani et al. \(2005\)](#) mentioned a dual state-parameter estimation of hydrological models through combining the EnKF with kernel smoothing technique. The kernel smoothing technique aimed to overcome drawbacks of parameter sampling in conventional model calibration methods through constructing artificial parameter evolution at each time step by adding small random perturbations ([Todini et al., 1976](#)). The drawback of such parameter sampling is the over-dispersion of parameter samples and posterior distribution of parameters is possibly too diffusive. The dual approach neglected effect of cross-state and parameter dependencies. In addition, there was no discussion concerning filter divergence caused by narrowing of parameter variance in Moradkhani et al.'s study. Hence in this study, we introduce a smoothed ensemble Kalman filter (SEnKF) to estimate simultaneously state variables and parameters of a forest carbon flux partition model. Although the SEnKF also combines the EnKF with kernel technique similar to [Moradkhani et al. \(2005\)](#), the SEnKF is a joint approach rather than dual one and addresses filter divergence issue in similar way to [Anderson and Anderson \(1999\)](#).

In Section 2, we describe the methodology of the SEnKF and builds up the mathematical formulation. Section 3 contains a brief outline of a carbon flux partition model and flux data at the AmeriFlux station in Howland, Maine, USA. The simulation results are displayed and analysed in Section 4. Discussion concerning state-parameter estimation is carried out in Section 5, and some conclusions are presented in Section 6.

2. SEnKF method

The SEnKF is a sequential data assimilation method with three components: (1) a dynamic model used to forecast system states as well as a parameter evolution process, (2) observation data and the relationship between the data and the model states, and (3) an assimilation scheme for model–data synthesis ([Evensen and Van Leeuwen, 2000](#); [Evensen, 2003](#); [Raupach et al., 2005](#)).

2.1. Dynamic model

A dynamic model can be expressed as one or more discrete-time nonlinear stochastic processes.

$$X^{k+1} = f(X^k, U^k, \theta^k) + \varepsilon^k \quad (1)$$

where k denotes the time step, X^k is a vector of random state variables or object variables (such as carbon flux or storage attributes), f is the model operator as a propagation of model state (such as rates of change of net carbon fluxes), U^k is a set of externally specified time-dependent forcing variables (such as meteorological variables and soil properties), θ^k is a set of model parameters or auxiliary variables (such as light use efficiency and partition ratios), and the noise term ε^k accounts for both imperfections in model formulation and random variability in forcing variables and parameters.

To extend the applicability of the EnKF to simultaneous state-parameter estimation, we need to build an evolution of the parameters similar to that of the state variables:

$$\theta^{k+1} = g(\theta^k) + \tau^k \quad (2)$$

where g is a transition operator (such as a linear function, $g(\theta^k) = \theta^k$), and τ is a random error term. We will discuss their definitions below. Now we define $Y^k = (X^k, \theta^k)^T$, $M = (f, g)^T$ and $q^k = (\varepsilon^k, \tau^k)^T$, where T denotes transposition. Then (1) and (2) are changed into a standard state model:

$$Y^{k+1} = M(Y^k, U^k) + q^k \quad (3)$$

2.2. Observation data

The observation (Z^k) is related to the system state, external forcing variables, and parameters through an expression of the form

$$Z^k = H(X^k, U^k, \theta^k) + \delta^k \quad (4)$$

or

$$Z^k = H(Y^k, U^k) + \delta^k \quad (5)$$

where the operator H specifies the deterministic relationship between the observation data and the model states. The noise term δ^k accounts for both measurement error (instrument and processing errors in the measurements) and representation error (errors in the model representation of Z , introduced by shortcomings in the observation model H), which is assumed to be Gaussian and independent of model error.

2.3. Assimilation scheme

The EnKF is based on the Monte Carlo method and the Kalman filter formulation to mimic the probability distribution of the model state, conditioned on a series of observations of the model state. The probability density of the model state is represented by a large ensemble of model states, and these are integrated forward in time by the model with a stochastic forcing term representing the model errors (Evensen, 1994). Each ensemble member evolves in time according to

$$Y_{j-}^{k+1} = M(Y_{j+}^k, U_j^k), \quad j = 1, \dots, N \quad (6)$$

where N denotes the number of model state ensemble members, Y_{j-}^{k+1} is the component of the j th ensemble member forecast at time $k+1$ and Y_{j+}^k is the j th updated ensemble member at time k . The noise term is not explicitly represented because the EnKF represents multiplicative model errors through forcing data perturbation (Evensen, 1997). The forcing data perturbations are made by adding white noise (subject to Gaussian distribution with zero mean and covariance Q_j^k) to forcing data at each time step:

$$U_j^k = U^k + \eta_j^k, \quad \eta_j^k = N(0, Q_j^k) \quad (7)$$

Now we discuss how to build an evolution of the parameters similar to that of the state variables. The conventional artificial parameter evolution, which adds a small random perturbation at each time step, results in over-dispersion of parameter samples and loss of continuity between two consecutive points in time. We used the kernel smoothing of parameter samples to remedy the problem, as described in West (1993):

$$\begin{aligned} \theta_{j-}^{k+1} &= a\theta_{j+}^k + (1-a)\bar{\theta}_+^k + \tau_j^k, \\ \tau_j^k &= N(0, h^2 V_+^k), \\ \bar{\theta}_+^k &= \text{mean}(\theta_{j+}^k), \\ V_+^k &= \text{var}(\theta_{j+}^k) \end{aligned} \quad (8)$$

where θ_{j-}^{k+1} is the component of the j th ensemble member forecast at time $k+1$, and θ_{j+}^k is the component of the j th updated ensemble member at time k , a is the shrinkage factor in $(0,1)$ of the kernel location, which is typically around 0.45–0.49, and h is the smoothing factor. In the study, determination of the smoothing factor is based on magnitude of ensemble variance V_+^k of parameters. Of course, criteria to determine the magnitude are somewhat subjective and generally depend on the background of the model. When V_+^k is quite large, h will be defined as $(1-a^2)^{1/2}$ and serve as variance reduction. While V_+^k progressively decreases and eventually may cause filter divergence, h ($h > 1$) needs to be chosen to inflate parameter spread. In general, the only viable method choosing h is experimentation (i.e., trial and error) to give an assimilation with the most favourable statistics (Anderson and Anderson, 1999; Annan et al., 2005).

Similarly, observation data are treated as random variables by generating an ensemble of observations from a distribution with the mean equal to the measurement value and a covariance equal to the estimated measurement error (Williams et al., 2005).

$$Z_j^{k+1} = Z^{k+1} + \delta_j^{k+1}, \quad \delta = N(0, R^{k+1}) \quad (9)$$

Because the true state is generally unknown, we calculate a forecasted ensemble covariance matrix to substitute for the definitions of the error covariance matrix in the Standard Kalman filter.

$$P_-^{k+1} = \frac{1}{N-1} [M_Y^{k+1} - \bar{M}_Y^{k+1}] [M_Y^{k+1} - \bar{M}_Y^{k+1}]^T \quad (10)$$

where $M_Y^k = [Y_{1-}^k - \bar{Y}_{1-}^k, \dots, Y_{N-}^k - \bar{Y}_{N-}^k]$ and $\bar{M}_Y^k = 1/N \sum_{j=1}^N Y_{j-}^k$. The updated scheme of the EnKF is as follows:

$$Y_{j+}^{k+1} = Y_{j-}^{k+1} + K^{k+1} (Z_j^{k+1} - H(Y_{j-}^{k+1}, U_j^{k+1})) \quad (11)$$

where K^{k+1} is Kalman gain.

$$K^{k+1} = P_-^{k+1} H^T (H P_-^{k+1} H^T + R^{k+1})^{-1} \quad (12)$$

3. Application of SEnKF to carbon (C) modelling

3.1. Flux partition model

We use a flux partition model (Reichstein et al., 2005) as our test dynamic model for the SEnKF method. Our selection is based on two considerations. First, it is an important model for constructing bottom-up estimates of continental carbon balance components. Second, it is appropriate for testing the robustness of the SEnKF method because it is nonlinear, there are sufficient observations of state variables, and it has multiple unknown parameters. The flux partition model divides net ecosystem exchange (NEE) into gross primary production (GPP) and total ecosystem respiration (RESP) as follows:

$$\begin{aligned} \text{NEE}_t &= \text{GPP}_t - \text{RESP}_t \\ \text{GPP}_t &= \text{LUE}_t \text{PAR}_t \text{NDVI}_t D_{\text{temp}} D_{\text{VPD}} \\ \text{RESP}_t &= R_{\text{ref},t} \exp \left[E_0 \left(\frac{1}{T_{\text{ref},t} - T_0} - \frac{1}{T_{\text{air},t} - T_0} \right) \right] \end{aligned} \quad (13a-c)$$

where subscript t denotes time-dependent, LUE_t is light use efficiency, PAR_t is photosynthetically active radiation, NDVI_t is the normalized difference vegetation index, $R_{\text{ref},t}$ is respiration when air temperature ($T_{\text{air},t}$) equals reference temperature ($T_{\text{ref},t}$, usually specified as 10°C), E_0 is temperature sensitivity, and T_0 is a datum of temperature to avoid a denominator of zero in the model (13c), kept constant at -46.02°C as

in Reichstein et al. (2005). D_{temp} determines the effect of temperature on photosynthesis, and D_{VPD} expresses the decrease in leaf exchange from both photosynthesis and transpiration caused by vapour pressure deficit (VPD), according to

$$D_{\text{temp}} = \max \left[\frac{(T_{\text{max}} - T_{\text{air}})(T_{\text{air}} - T_{\text{min}})}{(T_{\text{max}} - T_{\text{air}})(T_{\text{air}} - T_{\text{min}}) + (T_{\text{opt}} - T_{\text{air}})^2}, 0 \right] \quad (14)$$

$$D_{\text{VPD}} = 0.5 \left[1 + \frac{1}{1 + v_0 \exp(v_1 \text{VPD})} \right] \quad (15)$$

where T_{min} , T_{opt} , and T_{max} denote minimum, optimal, and maximum temperatures for photosynthesis, respectively, VPD is vapour pressure deficit, and v_0 and v_1 are two unknown coefficients. If we define state and driving force vectors as $Y_t = (\text{NEE}_t, \text{GPP}_t, \text{RESP}_t, \text{LUE}_t, T_{\text{min}}, T_{\text{opt}}, T_{\text{max}}, v_0, v_1, R_{\text{ref}}, E_0)$ and $U_t = (T_t, \text{PAR}_t, \text{VPD}_t, \text{NDVI}_t)$, then the model can be expressed in the form of (6).

3.2. Flux data

Eddy flux estimates of net ecosystem exchange (NEE) are based on the covariance of high frequency fluctuations in vertical wind velocity and CO_2 concentration (Baldocchi et al., 1988). We applied the SEnKF approach to data obtained at the AmeriFlux station in Howland, Maine, USA (Hollinger et al., 2004). The period was from 2000 to 2004 because there were sufficient hourly and daily data at the station for that time, and because NDVI data from the Moderate Resolution Imaging

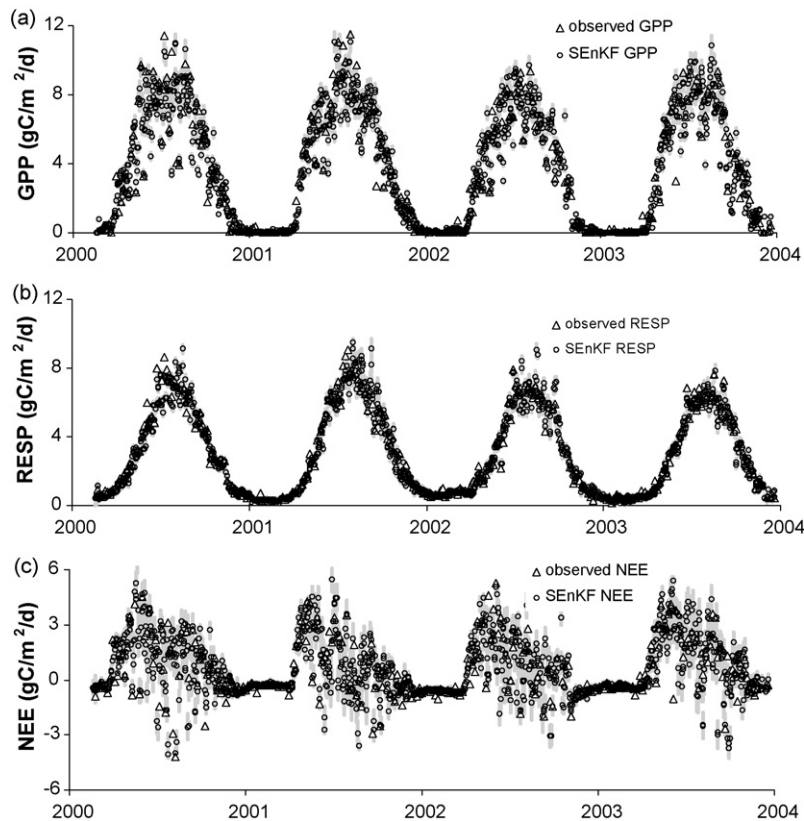


Fig. 1 – Time series of GPP, RESP and NEE simulated by the SEnKF and assimilated data. The grey vertical lines indicate the corresponding standard deviation around the mean of ensembles.

Spectroradiometer (MODIS) were not available until 2000. Field observations included hourly observations of NEE, humidity, photosynthetically active radiation (PAR), air temperature, air pressure, wind speed, and daily precipitation data. In our analysis, we used daily data of three state variables (NEE, GPP, and RESP) and four driving force variables (air temperature, PAR, VPD, and NDVI). NEE data were directly downloaded from the AmeriFlux Web station, in which NEE daily data were actually a composite of half-hour observations. RESP data were calculated from the temperature dependence curve of ecosystem respiration derived from nighttime NEE observations (Yuan et al., 2007). GPP data were pseudo-observations calculated as a total of NEE and RESP (13a). Gaps in carbon exchange data were filled using empirical regressions based on photosynthetically active radiation (PAR), air and soil temperature. And gaps in environmental factor data (e.g., VPD) were also filled using empirical regressions based on air temperature, relative humidity and saturation vapour pressure. The total percentage of 30-min periods that required gap filling over 4-year data

set was 15% for carbon exchange flux data and 12% for environmental factor data. Daily NDVI was calculated using linear interpolation of the MODIS 16-day composites. We assumed that data errors were subject to a Gaussian distribution with a zero mean and a variance of 20% of the average data based on uncertainty analysis in eddy covariance measurements (Hollinger and Richardson, 2005). The transition operator (H) in (4) was taken as a $3 \times N$ linear matrix with elements of 1 at diagonal nodes and 0 at other nodes.

To test the predictive power of the SEnKF, we held 80% of the data for model validation. Data assimilation was performed on only 20% of the observations.

4. Results and analysis

For 2000 to 2003, the total assimilated GPP, RESP and NEE were 4957 ± 21.7 , 4142 ± 16.8 , and $-815 \pm 22.3 \text{ gC m}^{-2}$, respectively. These compared well with observations for GPP, RESP,

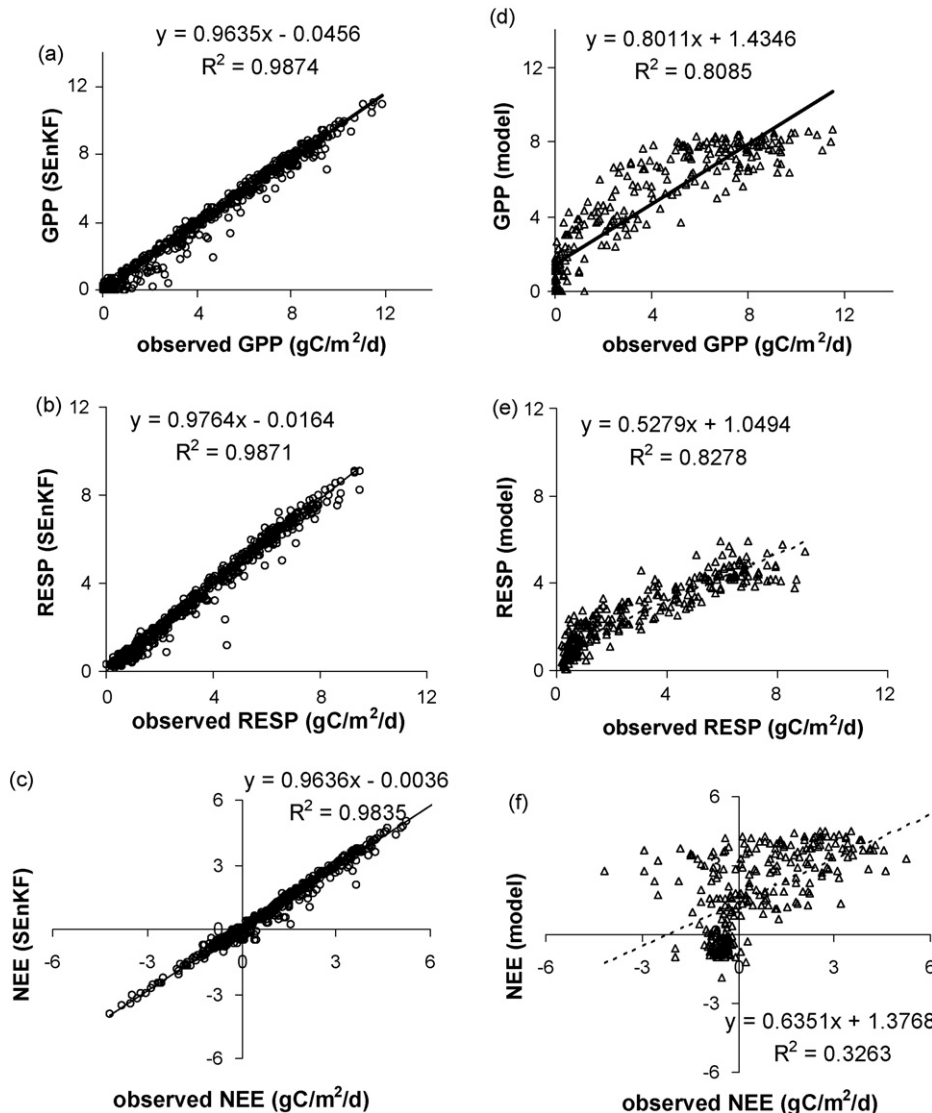


Fig. 2 – Comparison of simulation results by SEnKF (left panel) after assimilating 20% of observation data of state variables and by CMA (right panel) to observations. Note that we only assimilated 20% observation data of state variables into the model and left 80% observation data as validation.

and NEE of 5008 ± 50.1 , 4187 ± 41.9 , and $-821 \pm 24.6 \text{ gC m}^{-2}$, respectively. The estimate of NEE clearly indicates that the system was acting as a C sink during this time period. The maximum daily root-mean square error of the ensemble mean of GPP, RESP and NEE were 0.27, 0.26, and $0.23 \text{ gC m}^{-2} \text{ day}^{-1}$, respectively, which are small compared with their respective maximum daily means. Data assimilation effectively captured the temporal changes of GPP, RESP, and NEE over the 4 years (Fig. 1).

We evaluated the performance of the SEnKF method, comparing results from the SEnKF with those generated by the conventional modelling approach (CMA). The CMA had the same flux partition model but a set of fixed optimal parameter values was obtained using the conventional nonlinear inversion procedures in the statistical analytical software (SAS). However, this set of parameter values was derived from a mixture of 15 AmeriFlux stations, covering various ecosystem types (Yuan et al., 2007). Therefore, these parameter values were not necessarily optimal for the Howland, Maine station in this study. From Figs. 2 and 3, we can see that CMA predictions deviate from the observations at the Howland site,

caused at least partially caused by the use of time-invariant parameter values. In this paper, we refer to the CMA model as the base model. No further optimization of the parameter values was performed specifically for Howland in order to see whether or not the SEnKF could modify the bias of the model stemming from parameter uncertainty.

Fig. 2 compares the results simulated by the SEnKF data assimilation (left panel) and by the base model (right panel). Only 20% of the observations were used during data assimilation. The assimilated fluxes were very similar to the observations as indicated by the closeness of the points along the 1:1 line, whereas the flux estimates generated by the base model contained systematic errors as indicated by the deviation of triangles from the 1:1 line. Data assimilation accounted for more than 98% of the variation in the observations of GPP, REPS, and NEE, whereas the base model only explained for 81, 83, and 33% of the variation for GPP, RESP, and NEE, respectively. All the linear regression equations between assimilated and observed GPP, RESP, and NEE indicated no significant bias ($\alpha < 0.05$), whereas the base model generated strong biases.

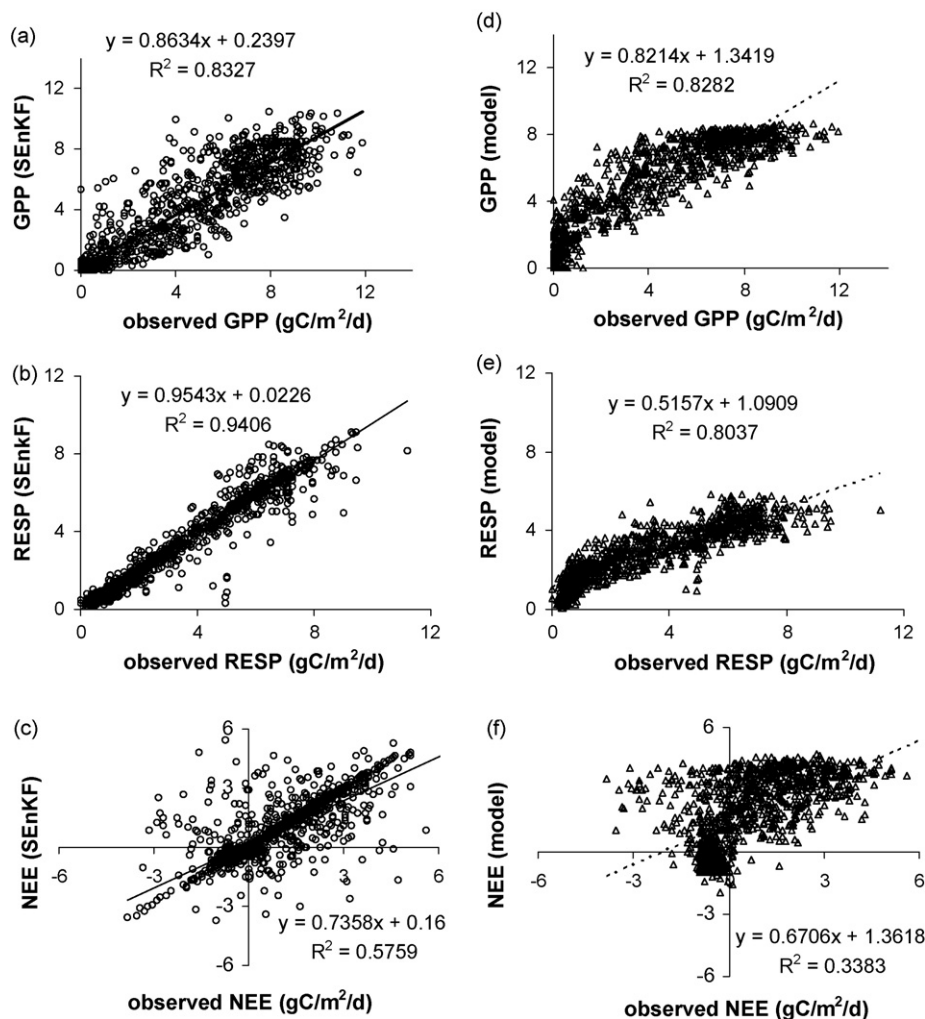


Fig. 3 – The left panels show the forecasted values of three state variables (GPP, RESP and NEE) of the model modified by the SEnKF against unassimilated data (80% of total observations). The right panels display the simulation results by the base model against the same unassimilated data.

The next step was to test whether the new parameter values derived from the SEnKF could be used to improve the prediction of system conditions. The left panel of Fig. 3 compares the model predictions (based on parameters derived from 20% of the observations using the SEnKF) with the observations that were held for validation (i.e., the 80% of the observations that the SEnKF did not see). The right panel of Fig. 3 compares the simulations generated by the base model to the same 80% of the observations. Comparing three pairs of corresponding linear fitting regression equations for GPP, RESP and NEE and their corresponding coefficients of determination (R^2), we see that the estimates of the three flux variables using the parameters (e.g., light use efficiency and reference respiration) modified by the SEnKF had less bias against observations than estimates using the base model (Fig. 3). This indicates the SEnKF can extend the model parameter values generated from data assimilation to predictions when observations are not available. This capability could be valuable for filling data gaps caused by instrument failure.

Predictions made by SEnKF with data assimilation matched observations substantially better (Fig. 2) than predictions made without data assimilation (the left panel of Fig. 3). The SEnKF procedure becomes regular Monte Carlo analysis at the time steps when no observation data are available for assimilation. The Monte Carlo analysis is based on Monte Carlo method and dynamic model under random perturbation of inputs and/or parameters in the specified error ranges to generate an ensemble of samples of state variables and then perform statistical analysis on the ensemble. At the data assimilation points (20% of the data), observation data strongly constrained the three flux variables (Fig. 2), and the differences between SEnKF model simulations and observations increased when no data were assimilated (left panels of Fig. 3). This might mean that prior states of the model have a weak effect on subsequent states. It might also suggest that there were other controlling factors (e.g., nutrition and forest age) affecting these processes which were not included in the model, or that uncertainty existed in observations. Fig. 4 compares standard ensemble variances by SEnKF data assimilation against those by Monte Carlo analysis without data assimilation. The SEnKF can reduce up to 70% of the variance of the ensemble without data assimilation. As the magnitude of the variance of the ensemble increases, the smoothing effect also increases. The variance measures the uncertainty stemming from parameters and driving forces.

Because the SEnKF can assimilate sequential observation data into the model, the SEnKF also revealed that the parameter values (e.g., light use efficiency and reference respiration) possessed strong seasonality or temporal variability (Fig. 5). This indicates that the base model has structural errors (manifested by nonoptimal parameter values in this study) and results in bias in prediction. The temporal change of parameter values was relatively smooth because a smoothing procedure was implemented in the SEnKF to control the over-dispersion of parameter sampling. In addition, the SEnKF can quickly stabilize the parameter values regardless of the initial values of the parameters (Fig. 6). These demonstrate that the SEnKF can be used to perform recursive model calibration to diagnose the adequacy of model structure.

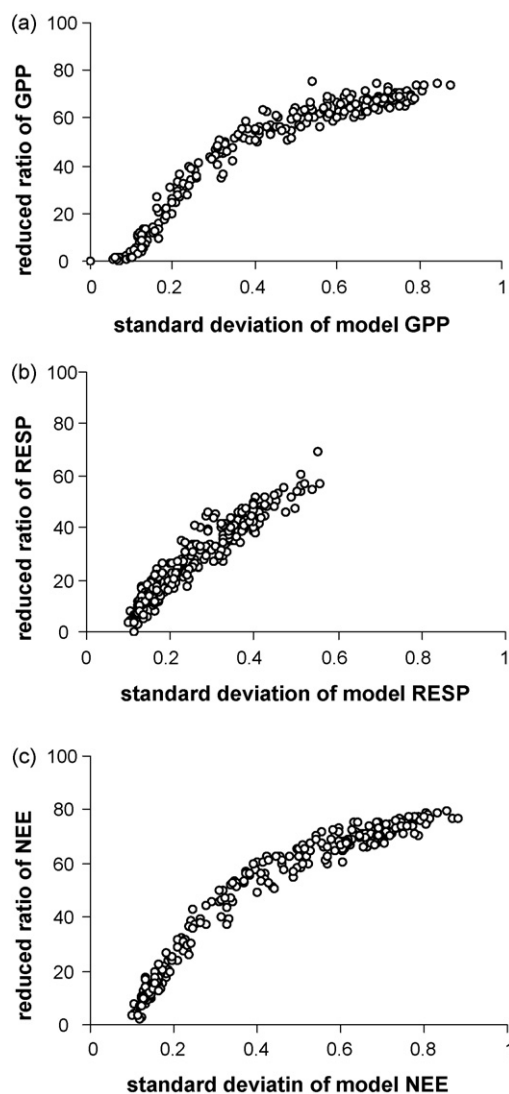


Fig. 4 – Reduced ratio of ensemble variances of GPP, RESP, and NEE generated by the SEnKF against that by the base model alone. The results show that the SEnKF can more dramatically reduce variances of state variables than the ensemble based only on Monte Carlo technique.

5. Comparison of state-parameter estimation methods based on the nonlinear filter

Here we only perform concept comparison of several nonlinear filters and their applications to state-parameter estimation rather than through numerical experiments.

5.1. Nonlinear filter

For nonlinear filtering problem expressed by (1), there has existed a complete theoretical resolution based on Bayesian principle (Anderson and Anderson, 1999). However, numerical implementation of the solution is difficult or even intractable. Hence there have developed many approaches to the problem, such as, Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Ensemble Kalman Filter (EnKF), Markov Chain Monte

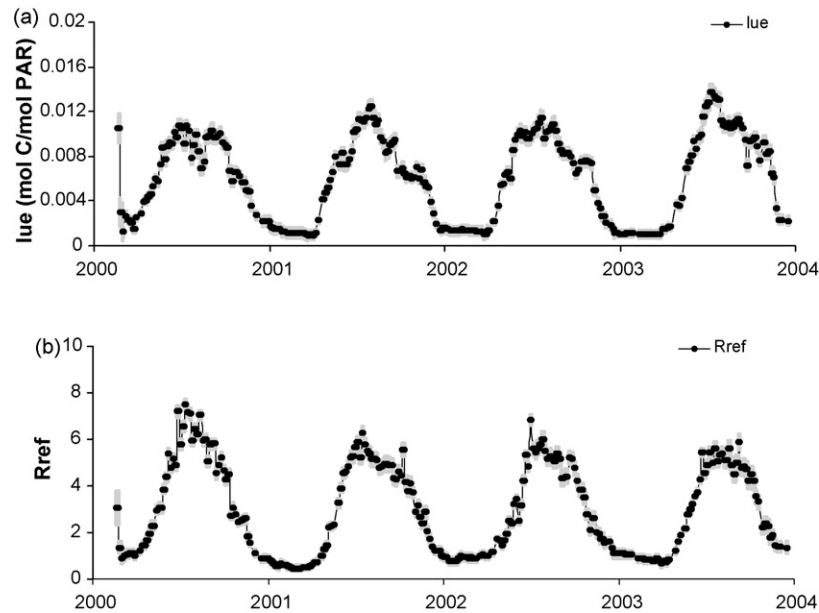


Fig. 5 – Temporal variations of two key parameters in the flux partition model: (a) light use efficiency (lue) and (b) reference respiration (Rref). The grey vertical lines indicate the standard deviation around the mean of ensembles.

Carlo (MCMC), and their variants. The EKF requires performing a truncated first-order Taylor linearization on the system equations about current state and then applying Kalman filter to the linearization system. Although the EKF has been used extensively, it suffers from possible divergence because of linearization for strongly nonlinear filter problems (Evensen, 1994). Furthermore, the EKF need a huge memory to keep covariance matrix in the integration forward process of state

variables. To address the shortcomings of the EKF, several new filtering methods such as UKF, EnKF and their variants have been introduced on the basis of Kalman filter. The common points of the new methods all adopt sampling approaches to conditional probability distribution rather than linearization of the nonlinear system. The main idea behind the UKF is to adopt a deterministic sampling of “sigma points” from the prior joint density and subsequently apply the nonlinear

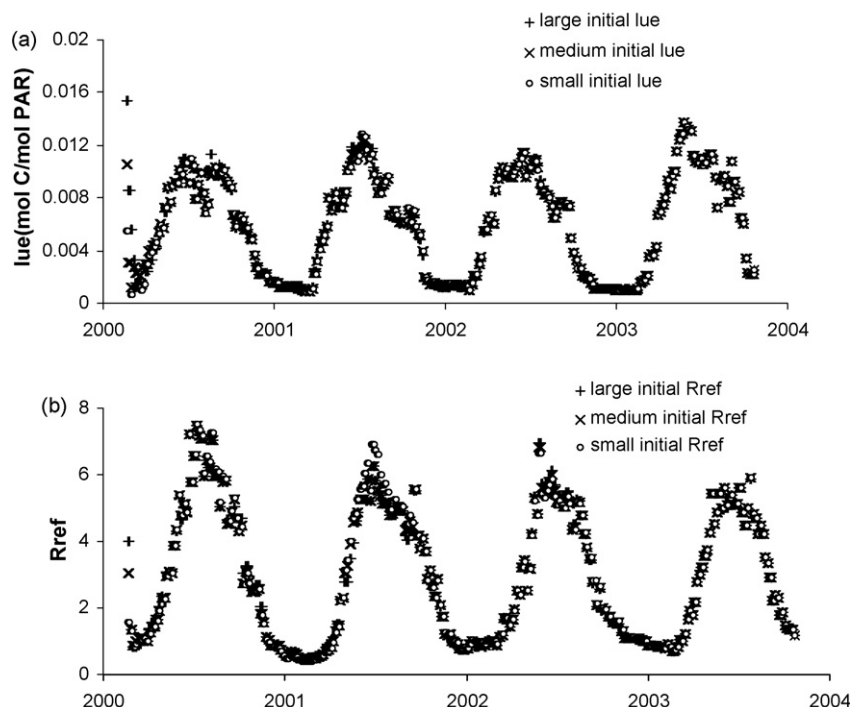


Fig. 6 – Stabilization of parameters by the SEnKF. Note the difference in the initial parameter values and the speedy convergence.

dynamics to these sampled points to obtain forecast values, finally update the forecast values through incorporating observations in the context of Kalman filter (Gove and Hollinger, 2006). Whereas the mechanism of the EnKF is to adopt Monte Carlo technique on the nonlinear dynamics to generate an ensemble of samples as a prior probability distribution and then incorporate the observations into the prior distribution according to formulation of Kalman filter to get a posterior probability distribution (Evensen, 1994, 2003). The UKF usually require that sample points are twice as many as dimension of state space. When the dimension of state space is very large, the computational burden of the UKF will be heavy. However, many data assimilations using EnKF have showed that the EnKF can well approximate the distribution with ensemble size being much less than dimension of state space (Aksoy et al., 2006; Moradkhani et al., 2005). The EnKF belongs to one kind of resampling Monte Carlo methods because its second step aims to update samples to generate by Monte Carlo in its first step. The art of MCMC is to set up a suitable Markov chain with a desired posterior distribution as stationary distribution and to judge when to stop simulation, i.e., to diagnose when the chain has practically converged. Several popular MCMC approaches include Gibbs sampling, Metropolis-Hastings and reversible jump (Arsham, 1998). The distinct advantage of these MCMC approaches is capable of representing probability distribution with non-Gaussian behaviour and higher-order moments of the distribution. However, their convergence speed is still a challenging problem.

5.2. Joint and dual estimation

Based on above nonlinear filters, there have been developed two principal approaches to simultaneously estimate state variables and parameters, namely joint and dual estimation. The joint approach is first to argument the state vector with parameters to form a joint state vector and then estimate the joint vector using a single filter recursion. The dual approach is to alternatively run two filters, one on the parameters in which the state are treated as known, and the other on the state vector in which the parameters are treated as known. The main difference between the two approaches, aside from the number of filters required, is that the joint filter explicitly allows for cross parameter and state dependencies. Whereas in the dual filtering approach the cross covariance are not explicitly estimated, so that it effectively assumes that the cross variances equal zero. Hence it could be argued that if correlation is suspected between states and parameters, the joint approach would be preferred (Gove and Hollinger, 2006). SEnKF in this study is a joint state-parameter approach. Whereas the state-parameter estimation developed by Moradkhani et al. (2005) is a dual approach although both SEnKF and Moradkhani's method are based on ensemble Kalman filter with kernel smoothing technique (West, 1993). The aim of applying a kernel-smoothing algorithm to an ensemble Kalman filter is to overcome the dramatic, sudden change of parameter values in time and the loss of information between two consecutive points in time. In addition, the kernel smoothing technique can also control narrowing of parameter variance because the narrowing usually results in filter divergence, through adjusting shrinking factor and smoothing factor in parameter

evolution (8). This is related to “conditional covariance inflation” method devised by Aksoy et al. (2006) and “covariance inflation” applied by Anderson (2001) and Annan et al. (2005). Above numerical experiment demonstrated that SEnKF can capture well time variability of parameters (Fig. 5) and quickly stabilize effect of initial state of parameters (Fig. 6). Recently several studies have also demonstrated the idea that carbon flux model parameters vary with time. For example, Gove and Hollinger (2006) used a dual unscented Kalman filter (UKF) to assimilate eddy flux data into a simple C cycle model to fill gaps in a flux record, estimating both parameters and states. They also found that model parameters varied through the season and ascribed that to deficiencies in their model.

6. Conclusions

The SEnKF method substantially and significantly improves flux estimates of a flux partition model and dramatically reduces uncertainties that stem from parameters and driving forces. Simultaneous parameter estimation can use near real-time observations to improve the predictive ability of dynamic models. The model based on the SEnKF can be used to fill data gaps in observations. This research demonstrates that the SEnKF is a robust and effective algorithm for evaluating and developing ecosystem models and improving understanding and quantification of model parameters and carbon cycle processes.

Acknowledgements

Work performed by M. Chen and S. Liu was performed under USGS contract 03CRCN0001. Wenping Yuan, from the Chinese Academy of Sciences, participated in the analysis of the flux data while he was visiting the U.S. Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS) in 2005. This research was funded by the Earth Surface Dynamics Program and the Geographic Analysis and Monitoring Program of the USGS.

REFERENCES

- Aksoy, A., Zhang, F., Nielsen-Gammon, J.W., 2006. Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea breeze model. *Mon. Weather Rev.* 134, 2951–2970.
- Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* 129, 2884–2903.
- Anderson, J.L., Anderson, S., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* 127, 2741–2758.
- Annan, J.D., Lunt, D.J., Hargreaves, J.C., Vades, P.J., 2005. Parameter estimation in an atmospheric GCM using the ensemble Kalman filter. *Nonlinear Process. Geophys.* 12, 363–371.
- Arsham, H., 1998. Techniques for Monte Carlo optimizing. *Monte Carlo Methods Appl.* 4 (3), 181–229.
- Baldocchi, D.D., Hicks, B.B., Meyers, T.P., 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology* 69 (5), 1331–1340.

- Bond-Lamberty, B., Wang, C., Gower, S.T., 2005. Spatiotemporal measurement and modeling of stand-level boreal forest soil temperatures. *Agric. For. Meteorol.* 131 (1–2), 27–40.
- Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.* 11 (2), 335–355.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* 99 (C5), 10143–10162.
- Evensen, G., 1997. Advanced data assimilation for strongly nonlinear dynamics. *Mon. Weather Rev.* 125 (6), 1342–1354.
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.* 53 (4), 343–367.
- Evensen, G., Van Leeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* 128 (6), 1852–1867.
- Gove, J.H., Hollinger, D.Y., 2006. Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *J. Geophys. Res.* 111, D08S07, doi:10.1029/2005JD006021.
- Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiol.* 25 (7), 873–885.
- Hollinger, D.Y., Aber, J., Dail, B., Davidson, E.A., Goltz, S.M., Hughes, H., Leclerc, M.Y., Lee, J.T., Richardson, A.D., Rodrigues, C., Scott, N.A., Achuatavari, D., Walsh, J., 2004. Spatial and temporal variability in forest-atmosphere CO₂ exchange. *Global Change Biol.* 10, 1689–1706.
- Knorr, W.G., Kattge, J., 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biol.* 11 (1), 1–19.
- Lorenz, E.N., 1996. Predictability: a problem partly solved. In: *Proceedings of the ECMWF seminar on predictability*, Reading, United Kingdom, ECMWF, pp. 1–18, vol. 1.
- Moradkhani, H., Sorooshian, S., Gupta, H.V., Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Adv. Water Resources* 28, 135–147.
- Nichol, C.J., Lloyd, J., Shibistova, O., Arneth, A., Roser, C., Knohl, A., Matsubara, S., Grace, J., 2002. Remote sensing of photosynthetic-light-use efficiency of a Siberian boreal forest. *Tellus Ser. B* 54 (5), 677.
- Raupach, M.R., Rayner, P.J., Barrett, D.J., DeFries, R.S., Heimann, M., Ojima, D.S., Quegan, S., Schimmlus, C.C., 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* 11 (3), 378–397.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránek, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biol.* 11 (9), 1424–1439.
- Todini, E., Szollosi-Nagy, A., Wood, E.F., 1976. Adaptive state-parameter estimation algorithm for real time hydrologic forecasting: a case study. In: *Proceedings of the IISA/WMO Workshop on the Recent Developments in Real Time Forecasting/Control of Water Resources Systems*, Luxemburg, Austria.
- Wang, Y.P., Baldocchi, D., Leuning, R., Falge, E., Vesala, T., 2006. Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. *Global Change Biol.* 12 (1), 1–19.
- West, M., 1993. Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Comput. Sci. Stat.* 24, 325–333.
- Williams, M., Schwarz, P.A., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biol.* 11 (1), 89–105.
- Yuan, W., Liu, S., Zhou, G.S., Zhou, G.Y., Tieszen, L.L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A.H., Goulden, M.L., Hollinger, D.Y., Hu, Y., Law, B.E., Stoy, P.C., Vesala, T., Wofsy, S.C., and other AmeriFlux collaborators, 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. *Agric. For. Meteorol.* 143, 189–207.